

Appln. No.: 10/524,965
Amendment Dated January 8, 2007
Reply to Office Action of November 8, 2006

RECEIVED
CENTRAL FAX CENTER

JAN 08 2007

Amendments to the Claims: This listing of claims will replace all prior versions, and listings, of claims in the application

Listing of Claims:

1. (Currently Amended) A method for generating structured document files from a document image, the method comprising the steps of:

segmenting the document image into ~~one or more~~ at least three zones, at least one of the ~~one or more~~ at least three zones containing a respective text image;

converting the respective text images within the at least one of the ~~one or more~~ at least three zones to digital text;

automatically identifying layout information for each of the ~~one or more~~ at least three zones;

comparing a layout graph of the document image with a layout graph of each model file of a plurality of model files, the layout graph of the document image representing a spatial relationship among the at least three zones in the document image and the layout graph of the respective model file representing an other spatial relationship among schema elements associated with the respective model file;

selecting a model file of a plurality of model files based on the comparison of the layout graph of the document file with the layout graph of each model file, each of the model files representing a respective type of document image, and including a respective schema for comparison with each of the ~~one or more~~ zones of the document image;

labeling each of the ~~one or more~~ at least three zones in accordance with the schema; and

automatically associating mark-up language tags with the labeled zones to generate the structured document files responsive to the identified layout information and the selected model file.

2. (Currently Amended) The method of claim 1, wherein the selected model file is associated with the schema and wherein the labeling step comprises at least the steps of:

automatically labeling each of the ~~one or more~~ at least three zones responsive to the selected model file.

Appn. No.: 10/524,965
Amendment Dated January 8, 2007
Reply to Office Action of November 8, 2006

3. (Currently Amended) The method of claim 1, further comprising the steps of:

receiving editing commands corresponding to the ~~one or more~~ at least three zones; and updating the ~~one or more~~ at least three zones responsive to the editing commands.

4. (Currently Amended) The method of claim 3, wherein the step of receiving editing commands includes the step of receiving text editing commands and the step of updating the ~~one or more~~ at least three zones includes the step of editing the digital text responsive to the text editing commands.

5. (Currently Amended) The method of claim 3, wherein the step of receiving editing commands includes the step of receiving segmenting commands and the step of updating the ~~one or more~~ at least three zones includes the step of updating characteristics of the ~~one or more~~ at least three zones responsive to the segmenting commands.

6. (Original) The method of claim 1, further comprising the step of: receiving editing commands corresponding to the schema; updating the schema responsive to the editing commands.

7. (Currently Amended) The method of claim 1, wherein the respective text images are displayed on a graphical user interface (GUI) and wherein the converting step comprises at least the step of:

overlays the respective text images displayed on the GUI with the at least one of the ~~one or more~~ at least three zones with the corresponding digital text.

8. (Presently Presented) The method of claim 1, wherein the structured document files include an XML file and an XSL file for each document image and wherein the generating steps comprises at least the step of:

formatting the XSL file such that information corresponding to each of the labeled zones in the XML file is displayed in multiple layers on a web browser.

Appn. No.: 10/524,965
Amendment Dated January 8, 2007
Reply to Office Action of November 8, 2006

9. (Original) The method of claim 1, wherein the steps of segmenting, converting, labeling, and automatically associating mark-up language tags are performed sequentially responsive to a selection of a workflow icon of a graphical user interface and wherein the method further comprises the step of:

updating the workflow icon to represent a next step of the segmenting, converting, labeling, and automatically associating mark-up language tags to be performed, wherein the workflow icon presents a unique image corresponding to each step.

10. (Currently Amended) A system for generating structured document files from a document image, the system comprising:

means for segmenting the document image into one or more at least three zones, at least one of the one or more at least three zones containing a respective text image;

means for converting the respective text images within the at least one of the one or more at least three zones to digital text;

means for automatically identifying layout information for each of the one or more at least three zones;

means for comparing a layout graph of the document image with a layout graph of each model file of a plurality of model files, the layout graph of the document image representing a spatial relationship among the at least three zones in the document image and the layout graph of the respective model file representing an other spatial relationship among schema elements associated with the respective model file;

means for selecting a model file of a plurality of model files based on the comparison of the layout graph of the document file with the layout graph of each model file, each of the model files representing a respective type of document image, and including a respective schema for comparison with each of the one or more zones of the document image;

means for labeling each of the one or more at least three zones in accordance with the schema; and

means for automatically associating mark-up language tags with the labeled zones to generate the structured document files responsive to the identified layout information and the selected model file.

Appn. No.: 10/524,965
Amendment Dated January 8, 2007
Reply to Office Action of November 8, 2006

11. (Currently Amended) The system of claim 10, further comprising:
means for receiving editing commands corresponding to the ~~one or more at~~
~~least three~~ zones; and
means for updating the ~~one or more at~~ ~~least three~~ zones responsive to the
editing commands.

12. (Currently Amended) The system of claim 10, further comprising:
means for receiving editing commands corresponding to the schema; and
means for updating the schema responsive to the editing commands.

13. (Currently Amended) A structured mark-up language generator for
generating structured document files from a document image, the generator comprising:
a document processor that: a) segments the document image into ~~one or~~
~~more at least three~~ zones, at least one of the ~~one or more at least three~~ zones
containing a respective text image, b) identifies layout information for each of the ~~one or~~
~~more at least three~~ zones, and c) converts the respective text images within the at least
one of the ~~one or more at least three~~ zones to digital text;
a model selector for: (1) comparing a layout graph of the document image
with a layout graph of each model file of a plurality of model files, the layout graph of the
document image representing a spatial relationship among the at least three zones in the
document image and the layout graph of the respective model file representing an other
spatial relationship among schema elements associated with the respective model file; and
(2) selecting a model file of a plurality of model files based on the comparison of the
layout graph of the document file with the layout graph of each model file, each of the
model files representing a respective type of document image, and including a respective
schema for comparison with each of the ~~one or more zones of the document image~~;
a labeler that labels each of the ~~one or more at least three~~ zones in
accordance with the schema; and
a structured document generator that generates the structured document
files responsive to the identified layout information and the selected model file.

14. (Currently Amended) The generator of claim 13, further comprising:

Appn. No.: 10/524,965
Amendment Dated January 8, 2007
Reply to Office Action of November 8, 2006

an editor coupled to the document processor that enables editing of the digital text and the one or more at least three zones.

15. (Currently Amended) The generator of claim 13, further comprising: an editor coupled to the labeler that enables editing of the labels for each of the one or more at least three zones.

16. (Currently Amended) A graphical user interface (GUI) for generating structured document files from a document image, the GUI comprising: a document panel for displaying a document image;

a schema panel for displaying a schema corresponding to the document image and a selected model file which is selected from a plurality of model files based on the comparison of the layout graph of the document file with the layout graph of each model file, each of the model files representing a respective type of document image, and wherein the layout graph of the document image representing a spatial relationship among zones in the document image and the layout graph of the respective model file representing an other spatial relationship among schema elements associated with the respective model file the selected model file includes the schema for comparison with each of the one or more zones of the document image; and

a workflow icon for directing the generation of at least one structured mark-up language document from the document image, the workflow icon reflecting a next step in a process to generate the at least one structured mark-up language document.

17. (Original) The GUI of claim 16, wherein the process includes sequentially performing the steps of loading an image, segmenting the image into zones, converting text within the zones to digital text, labeling the zones, and generating the at least one structured document and wherein the workflow icon is updated during the process to present unique images corresponding to each step to be performed in the process.

18. (Currently Amended) A computer readable medium including software that is configured to control a general purpose computer to implement a method for generating structured document files from a document image, the method comprising the steps of:

MATT-218US

Appln. No.: 10/524,965

Amendment Dated January 8, 2007

Reply to Office Action of November 8, 2006

segmenting the document image into one or more at least three zones, at least one of the one or more at least three zones containing a respective text image; converting the respective text images within the at least one of the one or more at least three zones to digital text;

comparing a layout graph of the document image with a layout graph of each model file of a plurality of model files, the layout graph of the document image representing a spatial relationship among the at least three zones in the document image and the layout graph of the respective model file representing an other spatial relationship among schema elements associated with the respective model file;

selecting a model file of a plurality of model files based on the comparison of the layout graph of the document file with the layout graph of each model file, each of the model files representing a respective type of document image, and including a respective schema for comparison with each of the one or more zones of the document image;

automatically identifying layout information for each of the one or more at least three zones;

labeling each of the one or more at least three zones in accordance with the schema; and

automatically associating mark-up language tags with the labeled zones to generate the structured document files responsive to the identified layout information and the selected model file.

19. (Currently Amended) The computer readable medium of claim 18, wherein the method implemented by the software configured general purpose computer further comprises:

updating the one or more at least three zones responsive to editing commands corresponding to the one or more at least three zones.

20. (Original) The computer readable medium of claim 18, wherein the method implemented by the software configured general purpose computer further comprises:

MATTI-218US

Appn. No.: 10/524,965

Amendment Dated January 8, 2007

Reply to Office Action of November 8, 2006

updating the schema responsive to editing commands corresponding to the schema.

21. (Previously Presented) The method of claim 6, further comprising the step of:

automatically updating the selected model file based on the updated schema.